Update on H->bb/cc/gg at 350 GeV

CLICdp Analysis Meeting - 6 June 2015

Marco Szalay

Max-Planck-Institut für Physik
(Werner-Heisenberg-Institut)

Excellence Cluster
Universe

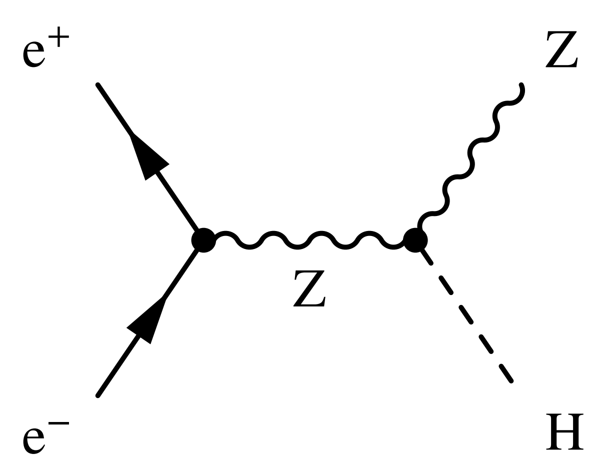
Outline

- Introduction
- Event selection and MVA performance
- Template Fit Status
- Background modelization
- Conclusion

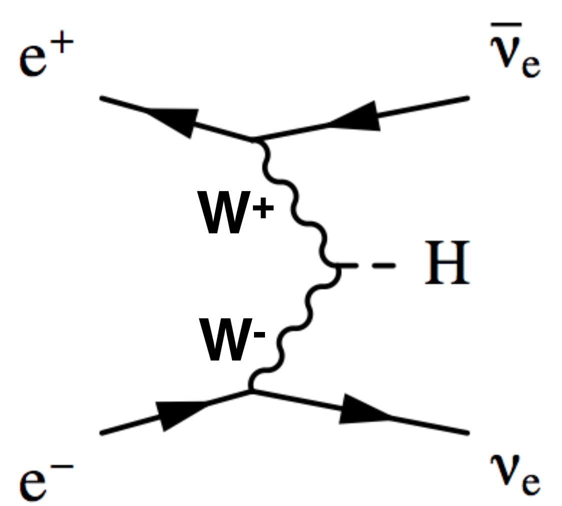
Introduction

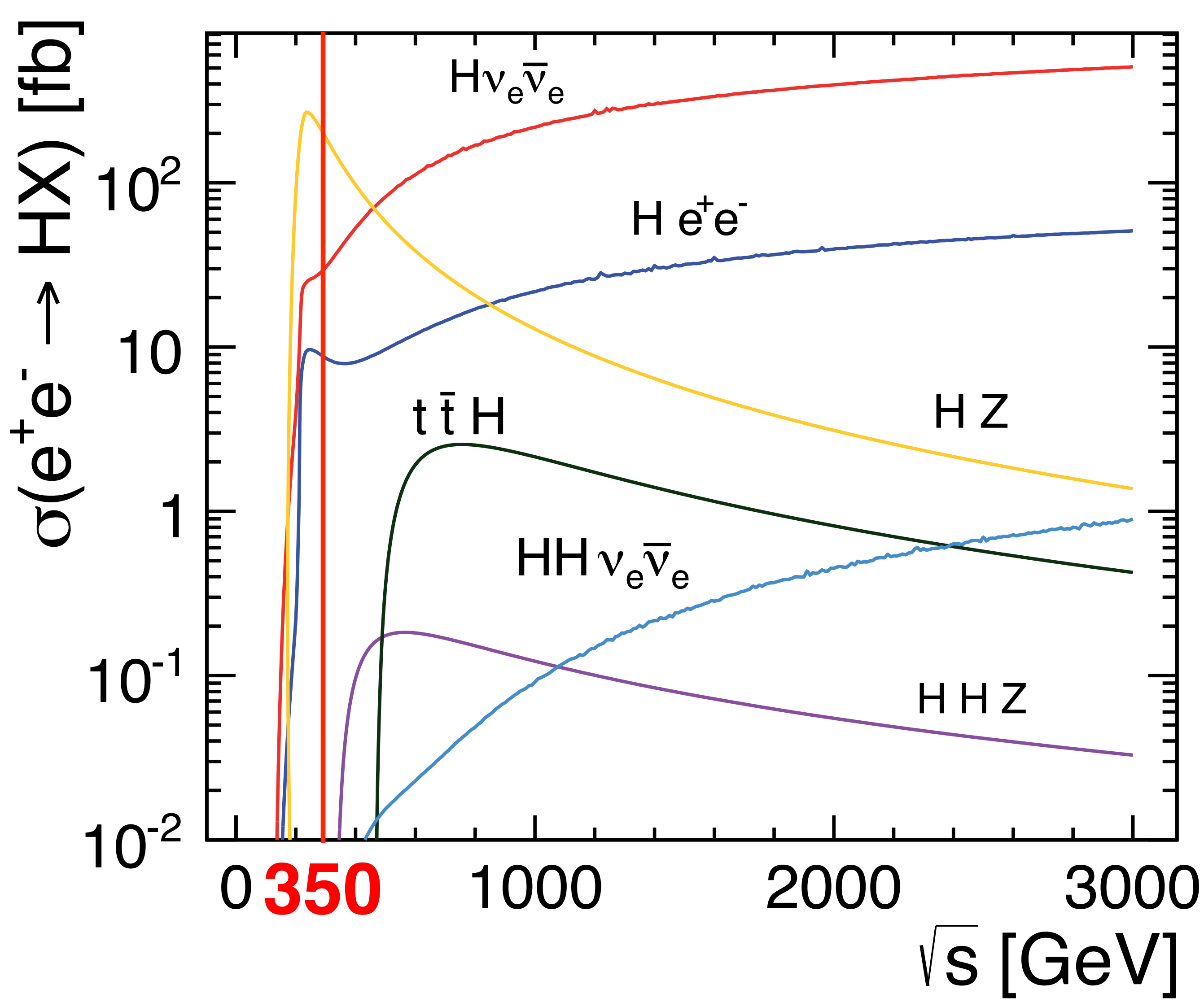
Main H production channels at 350 GeV:

- Higgs strahlung

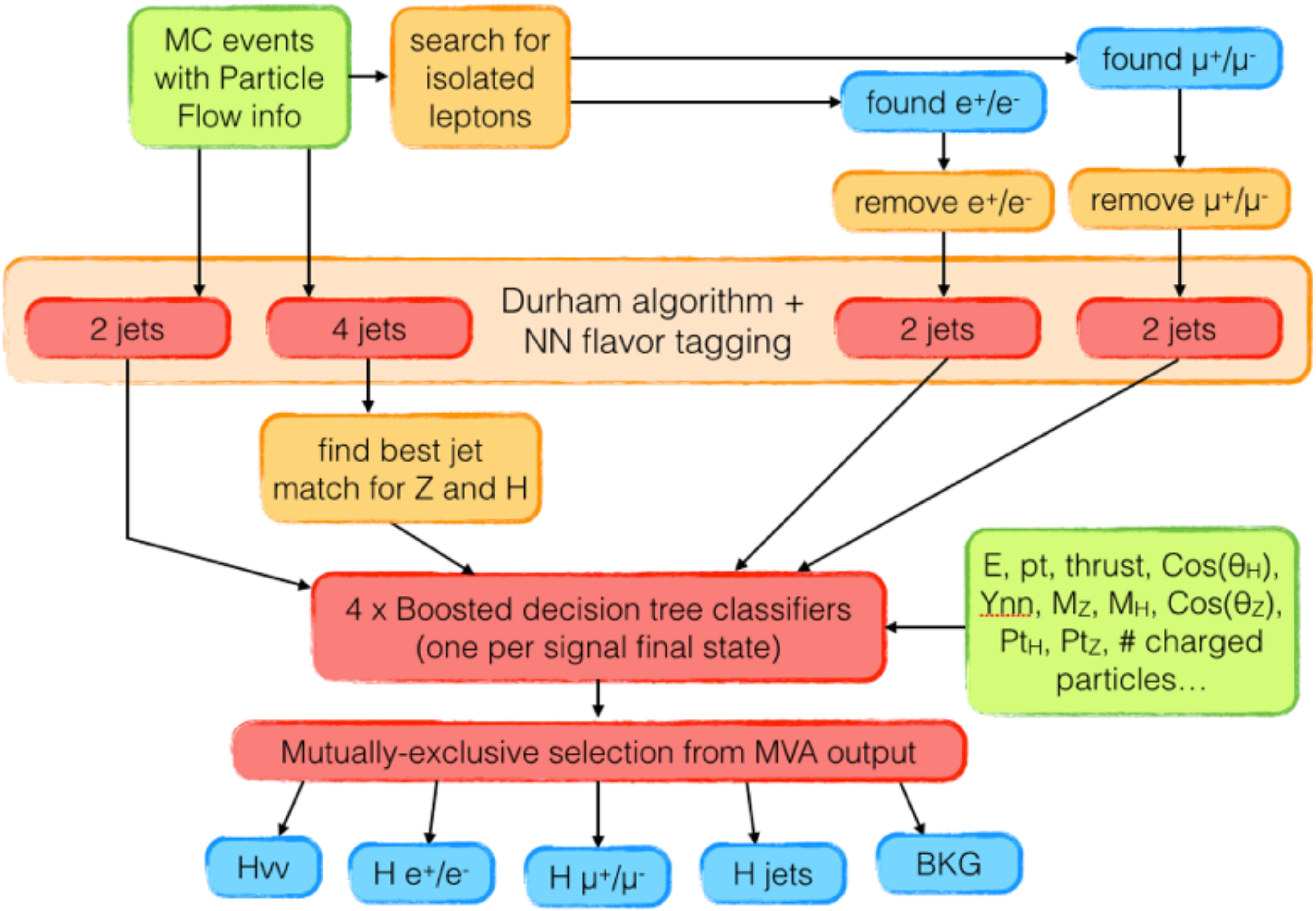


- Vector Boson Fusion





Event Selection

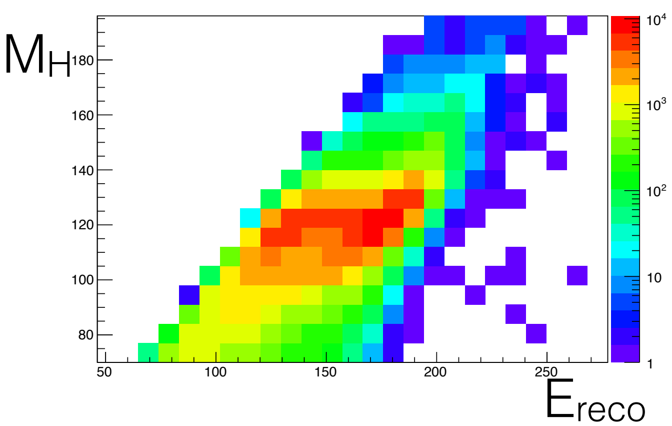


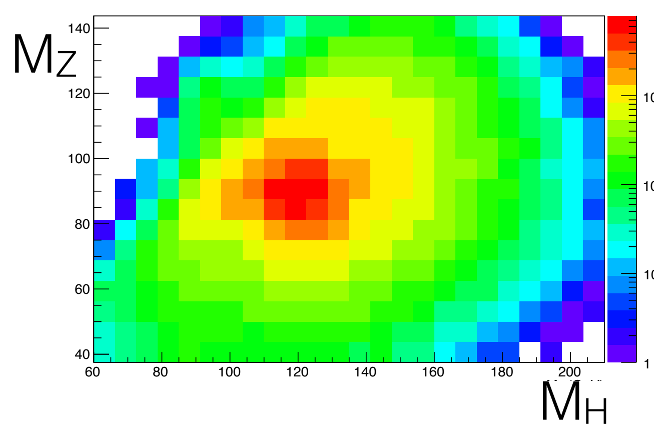
BDT Performance

Preselection:

Hνν
- $80 < M_H < 180$
- $60 < E_{reco} < 260$

HZ: Z→jets
- $70 < M_H < 200$
- $50 < M_Z < 130$

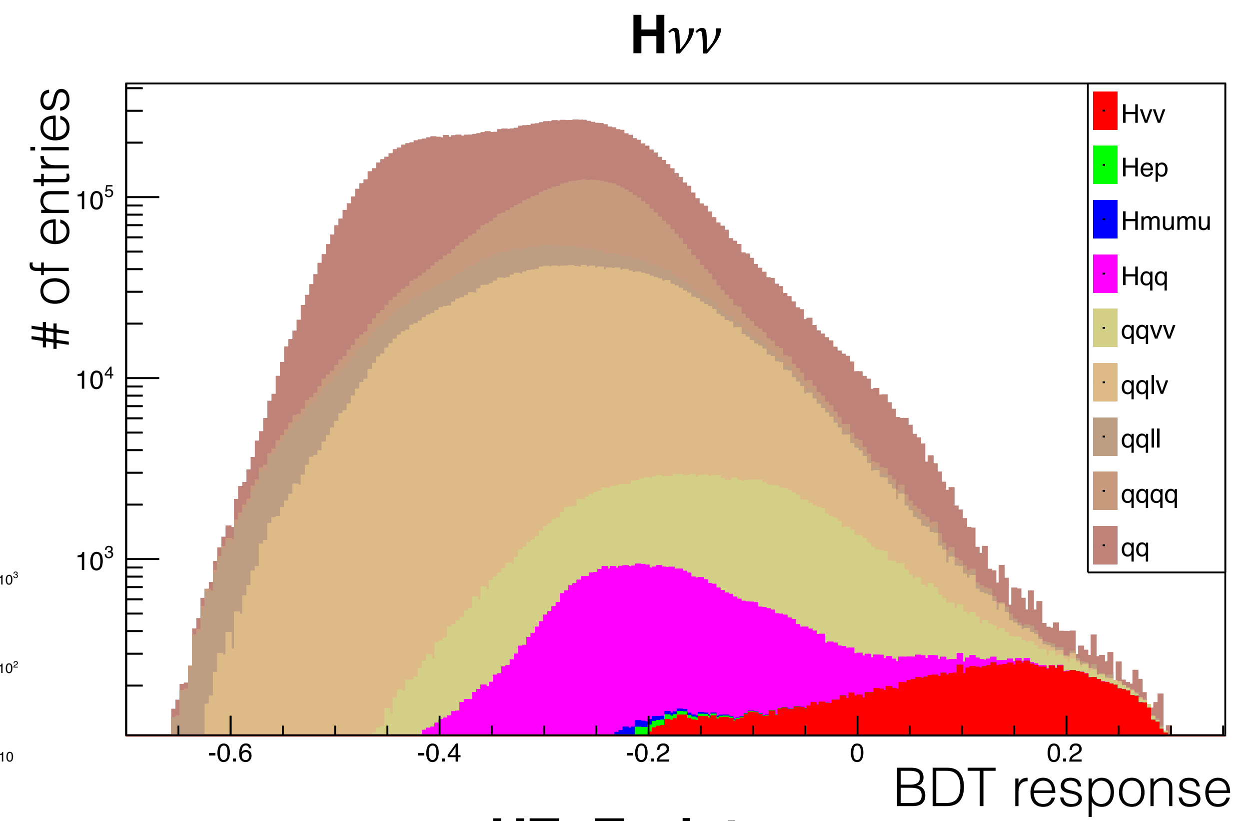


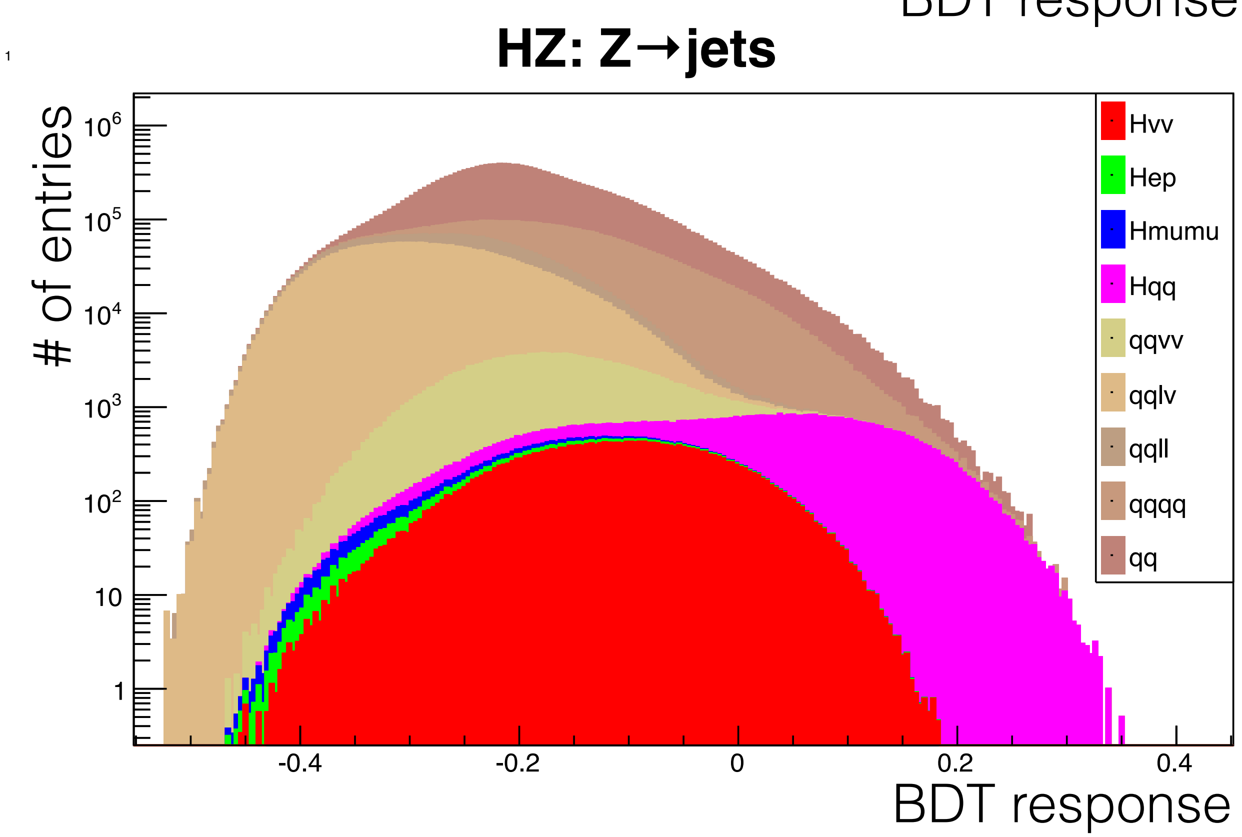


BDT ranking:

Hνν
1) M_H,
2) thrust,
3) E_{reco},
4) θ_H,
5) P_T,
6) ϕ_H

HZ: Z→jets
1) N_P
2) P_{max}
3) thrust
4) Cos θ_{thr}
5) θ_{ij}
6) M_Z





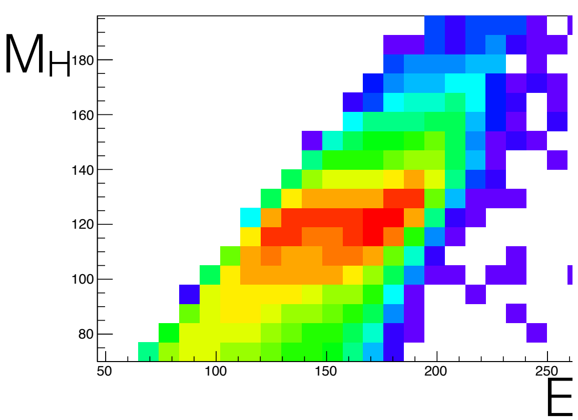
BDT Performance

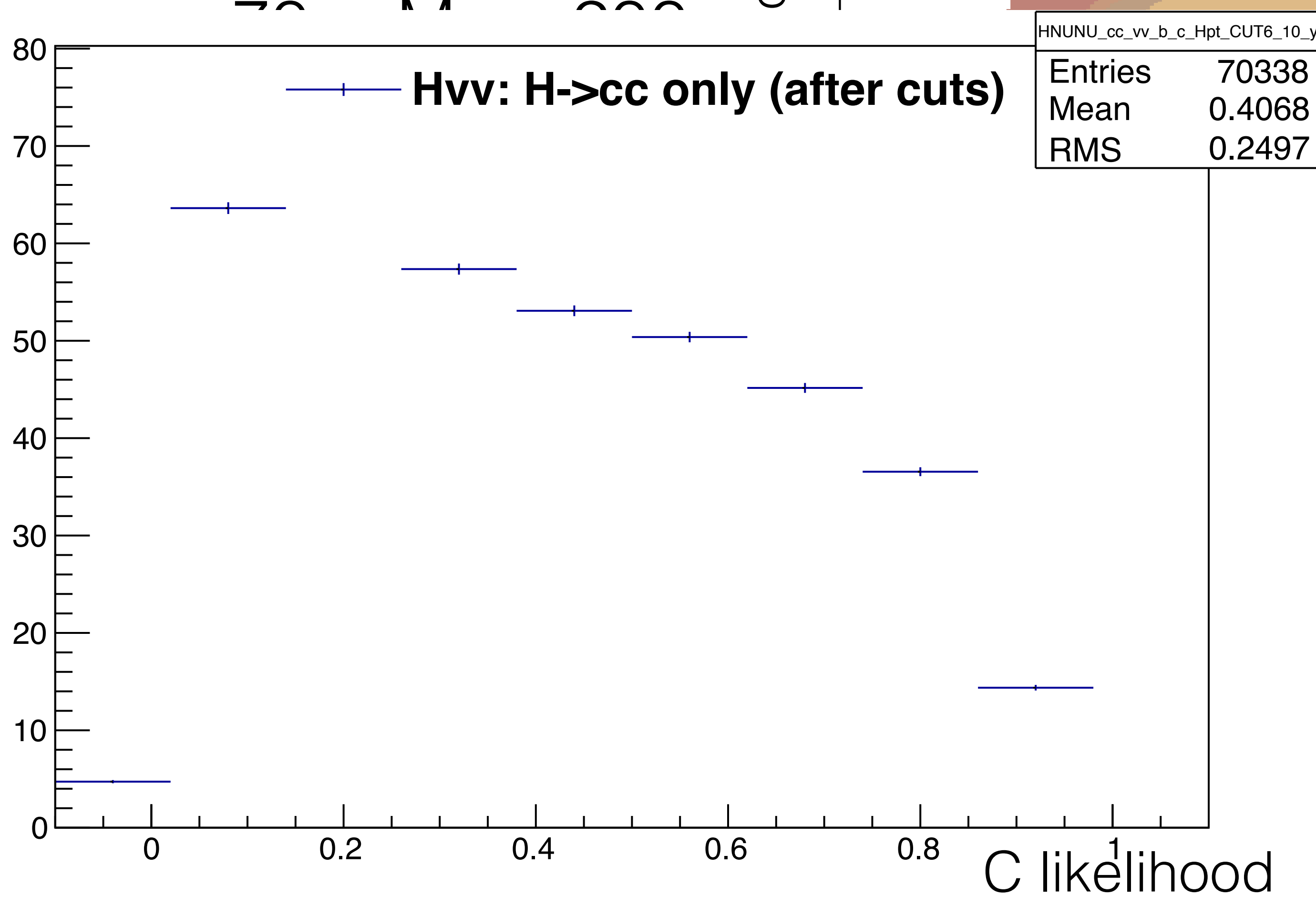
Preselection:

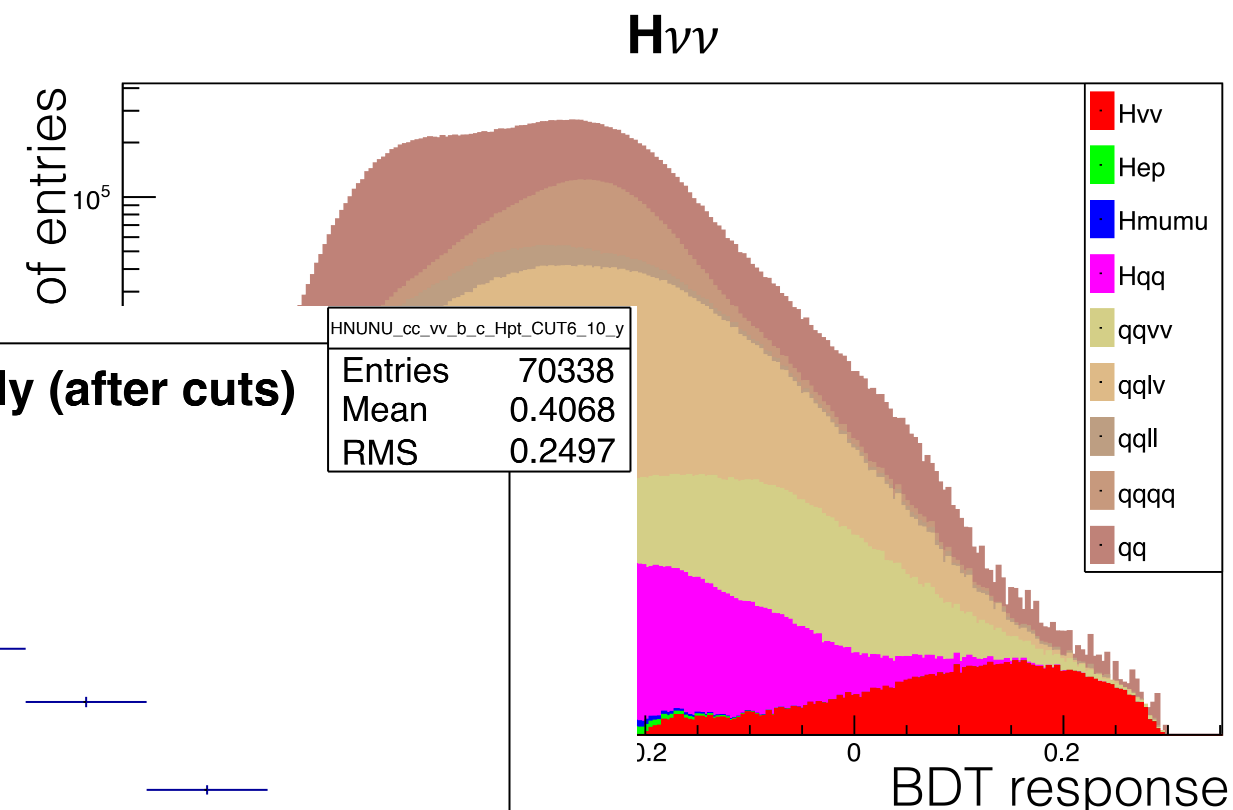
$H\nu\nu$

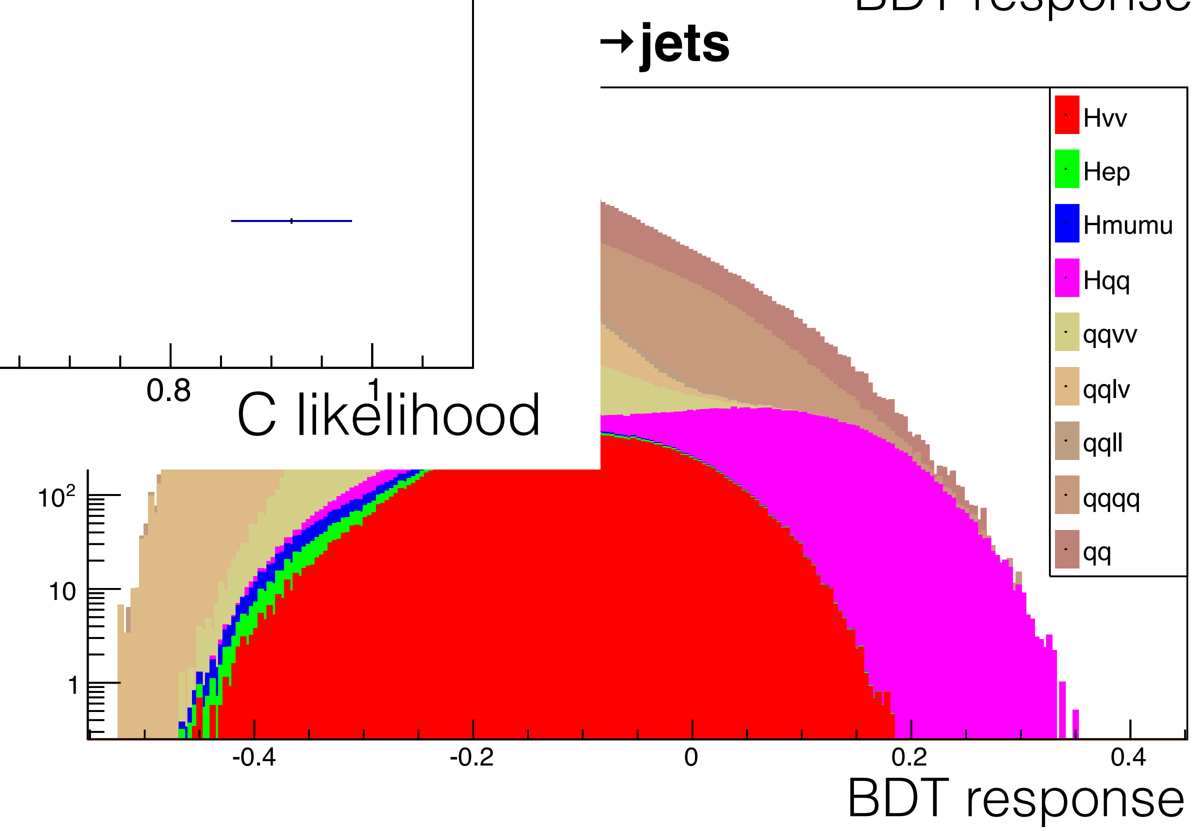
- $80 < M_H < 18$
- $60 < E_{reco} < 2$

HZ: Z→jets









BD

1) M_H,
2) thrust,
3) E_{reco},
4) $\boldsymbol{\theta}_H$,
5) P_T,
6) ϕ_H

3) thrust
4) Cos $\boldsymbol{\theta}_{thr}$
5) $\boldsymbol{\theta}_{ij}$
6) M_Z

Efficiency Improvements

Hvv cutflow	bb	cc	gg	qq	qqlv	qqvv
nocuts	14456	729	2185	1.2×10^7	2.95×10^6	162300
$60 < E_{Reco} < 260$	14449	728	2183	9.1×10^6	1.6×10^6	158484
$80 < H_M < 180$	14174	714	2147	6.9×10^6	741522	129504
BDT (Hqq) < 0.07	14173	714	2147	6.9×10^6	741522	129504
BDT (Hvv) > 0.11	8773	401	1071	4621	1975	1843
Total Efficiency	61%	55%	50%	3.7×10^{-4}	6.6×10^{-4}	1.1%

Hqq cutflow	bb	cc	gg	qq	qqqq
nocuts	26209	1322	3961	1.2×10^7	2.8×10^6
$50 < Z_M < 130$	25621	1294	3869	1.5×10^6	2.7×10^6
$70 < H_M < 200$	25615	1293	3869	1.37×10^6	2.7×10^6
BDT (Hvv) < 0.08	25338	1288	3859	1.37×10^6	2.7×10^6
BDT (Hqq) > 0.11	13434	566	1630	49994	2912
Total Efficiency	51%	43%	41%	3.9×10^{-3}	0.1%

Template Fit

- Binned maximum likelihood fit on multi-dimensional space: b and c likelihoods and H_{Pt}
- Assume Poissonian fluctuation for each data bin:

$$P_{ijk} = \frac{\mu^n e^{-\mu}}{n!}$$

with n = number of data entries in bin ijk
and $\mu = \sum w_m T_m$ for the same bin

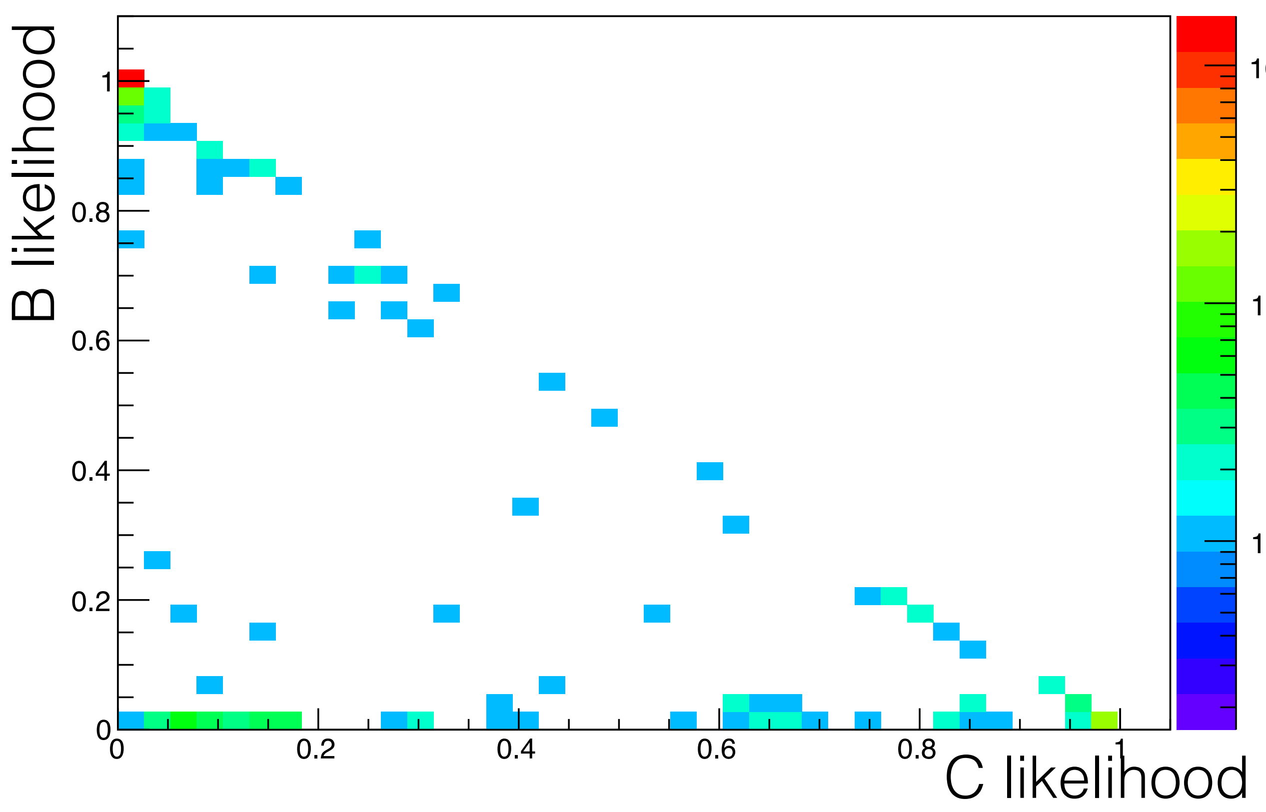
- Then the Likelihood is the product of P_{ijk} in all bins
- Find the w_m that maximize this value

Fit Methods

1. Log likelihood in 3D space (b-like, c-like and Hpt)
2. Log likelihood in multiple 1D projections
3. MCMC in multiple 1D projections (BAT toolkit)

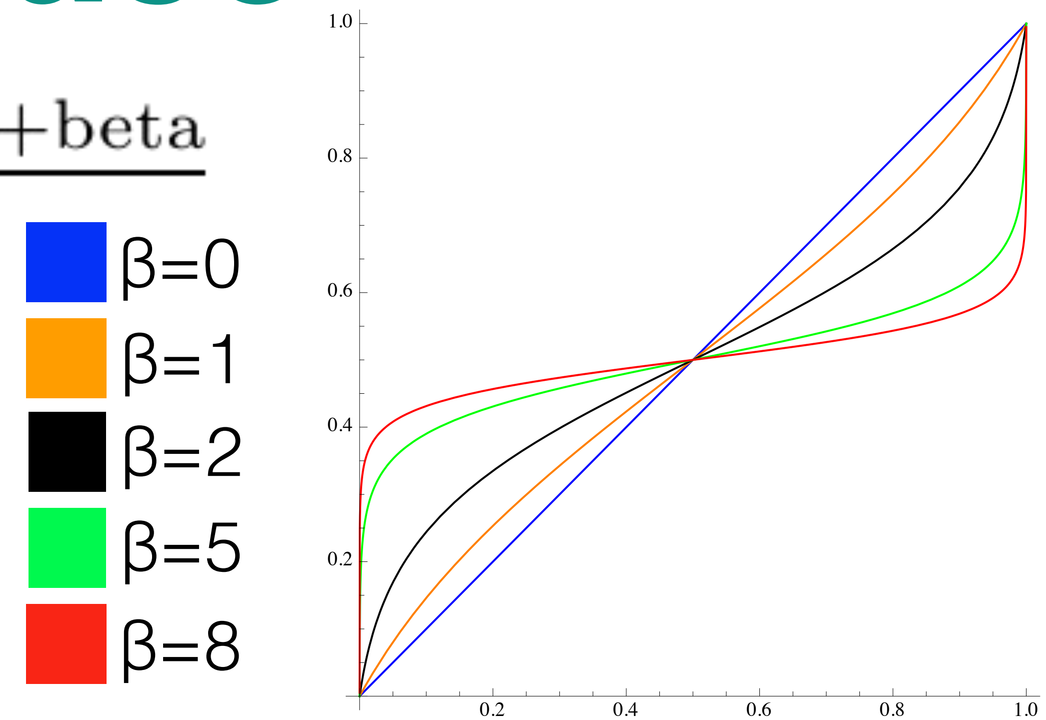
- All give similar results for coarse binning (~10-12% in Hvv:H->cc), only 3D Log likelihood gets better with finer binning, probably due to template artifacts

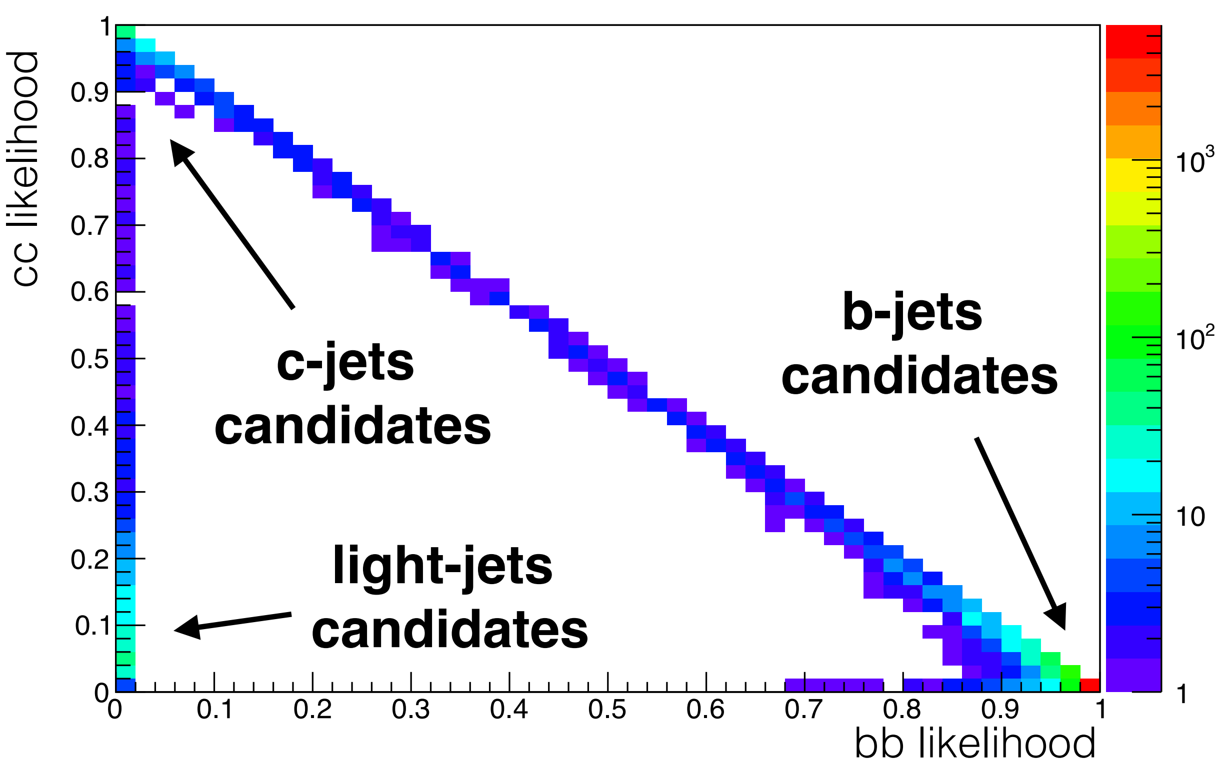
ee->qq background sample, unnormalized



Remapping the Flavor Space

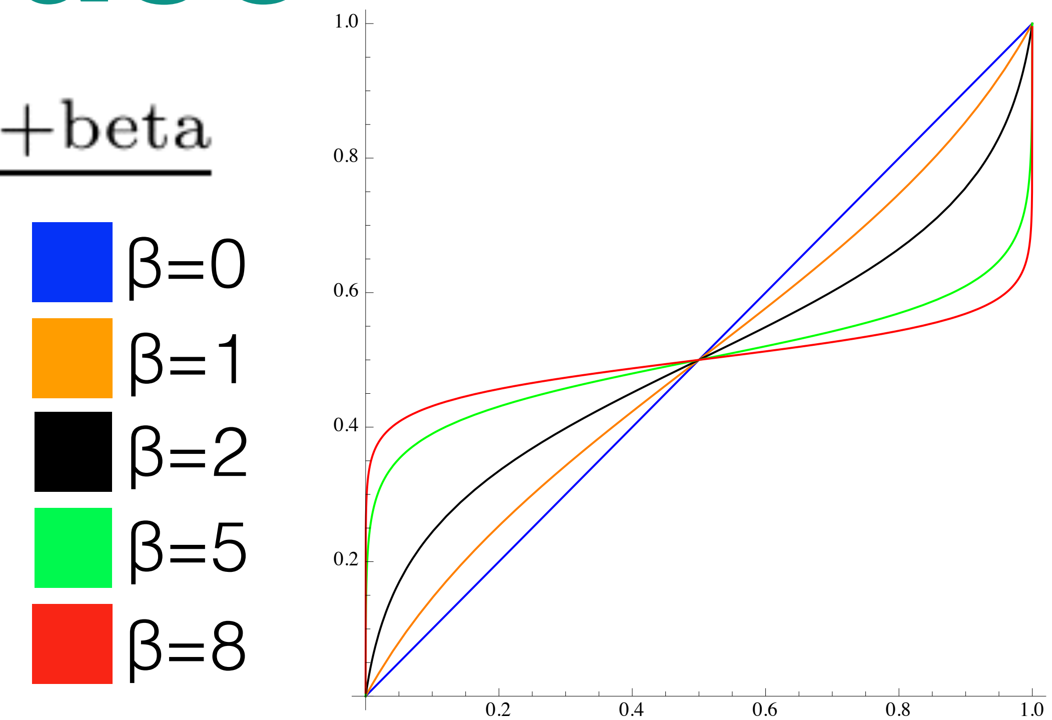
$$f(x, beta) = \frac{\tanh^{-1}((2x-1)\tanh(\mathrm{beta}))+\mathrm{beta}}{2\mathrm{beta}}$$

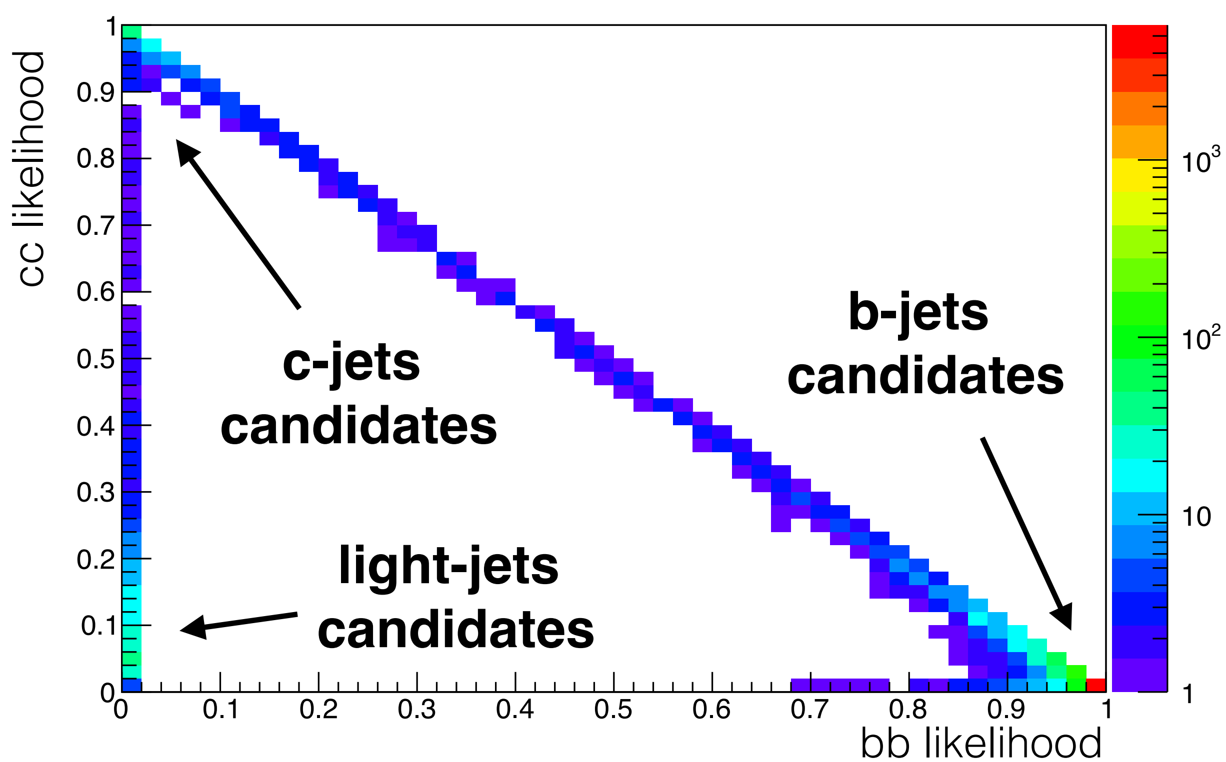


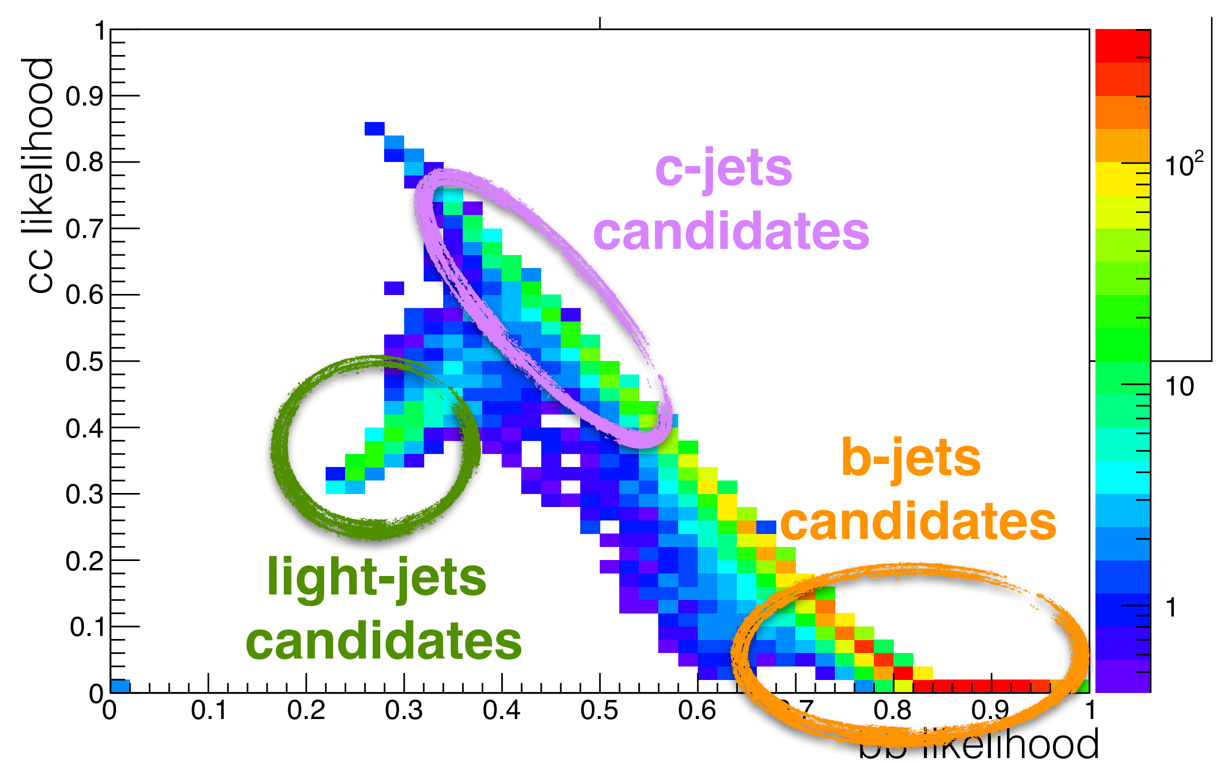


Remapping the Flavor Space

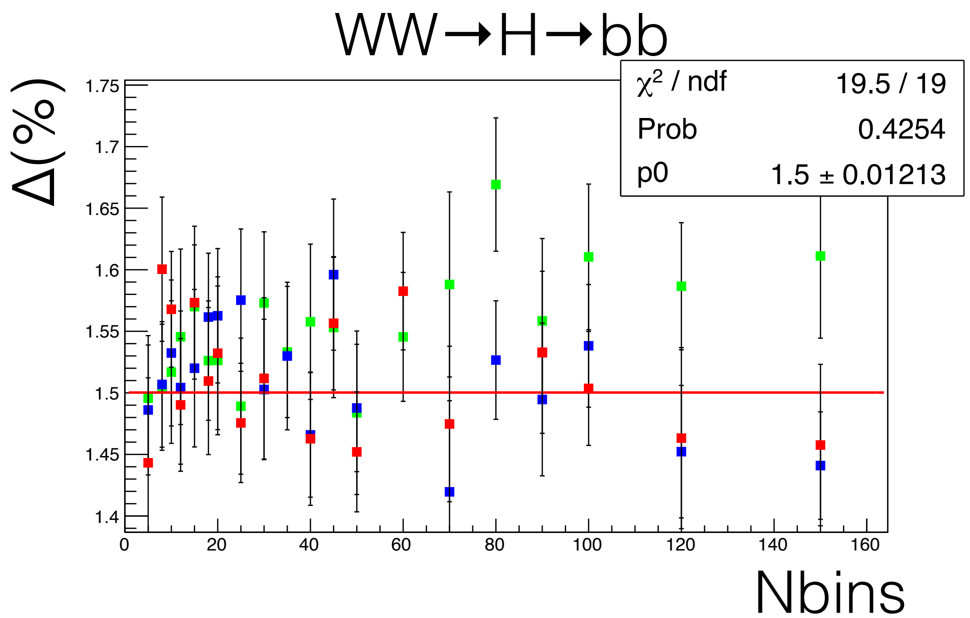
$$\boldsymbol{f}(x,beta) = \frac{\tanh^{-1}((2x-1)\tanh(\text{beta}))+\text{beta}}{2\text{beta}}$$

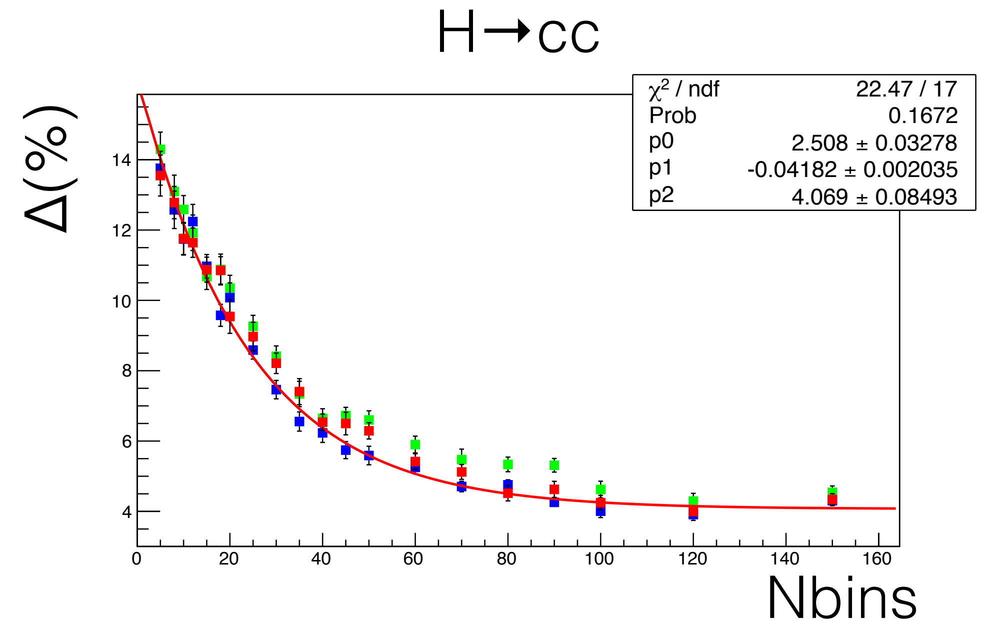


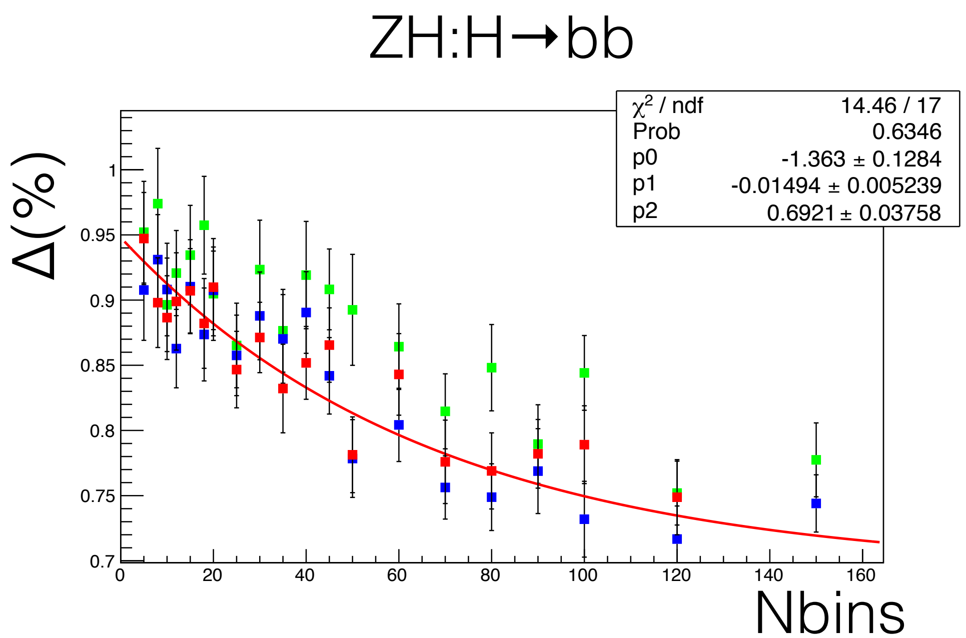


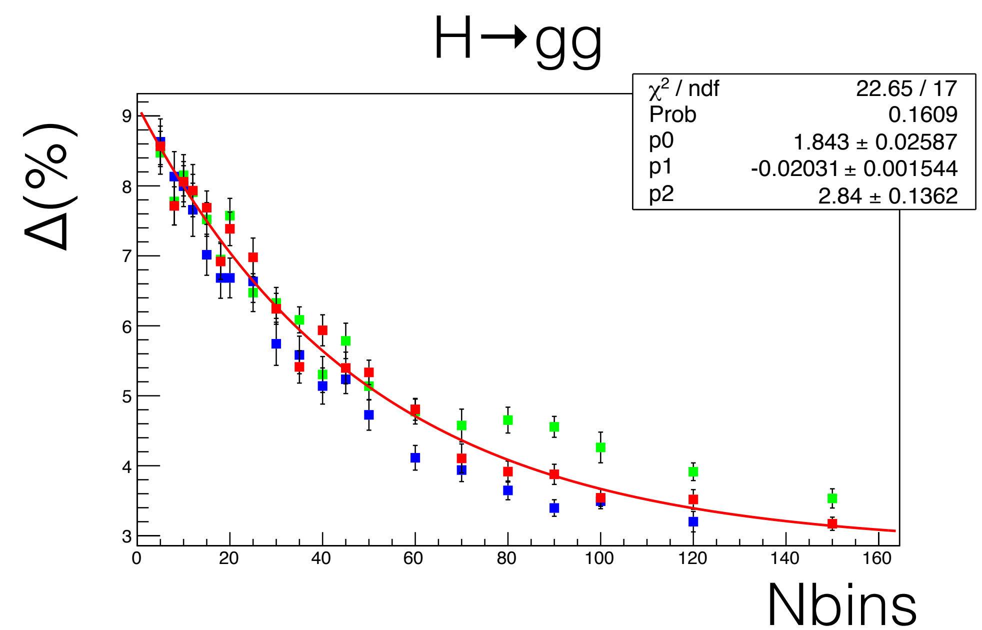


Remapping Results

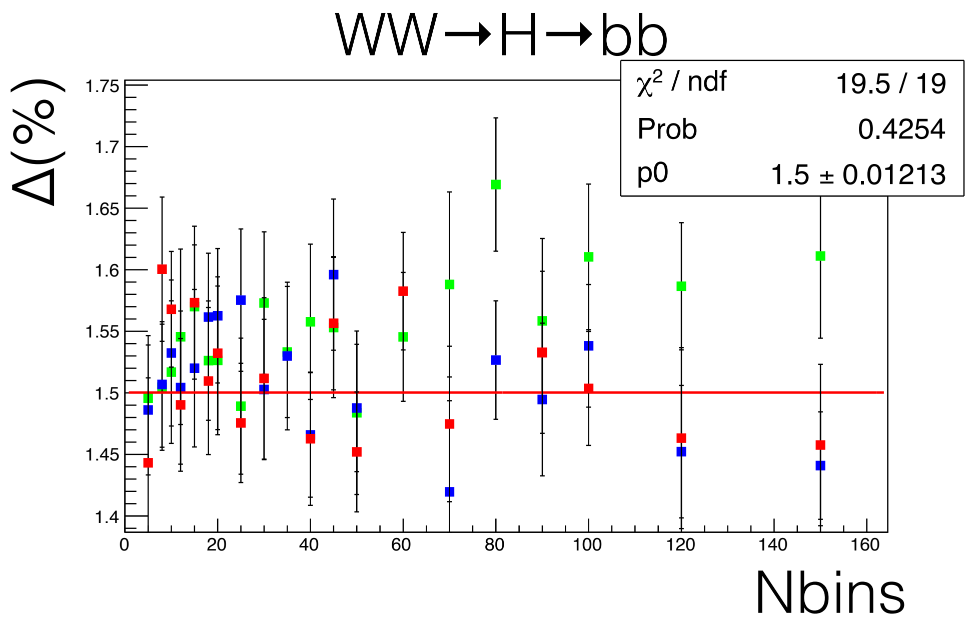


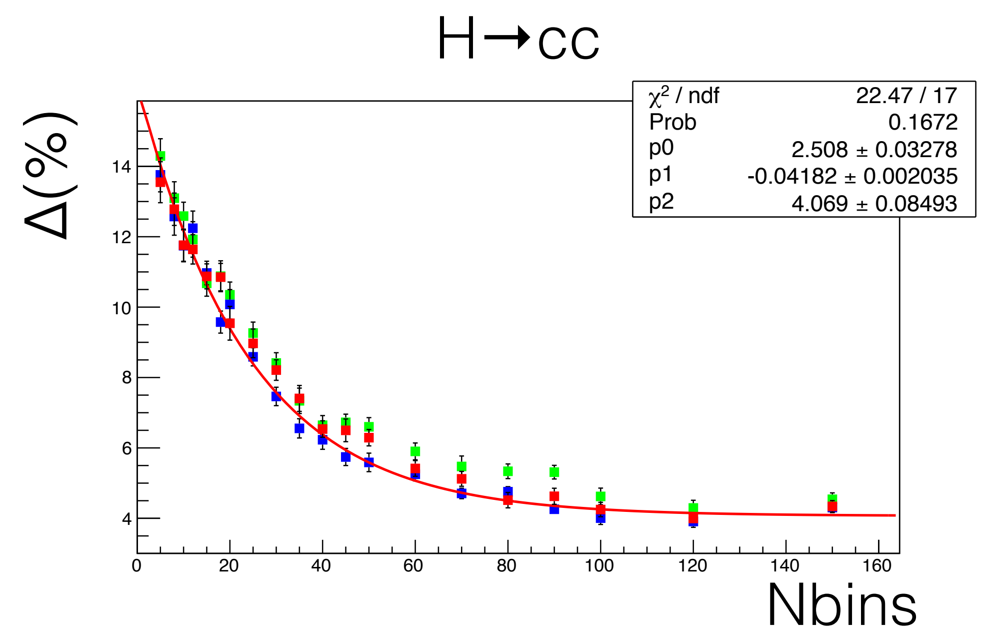


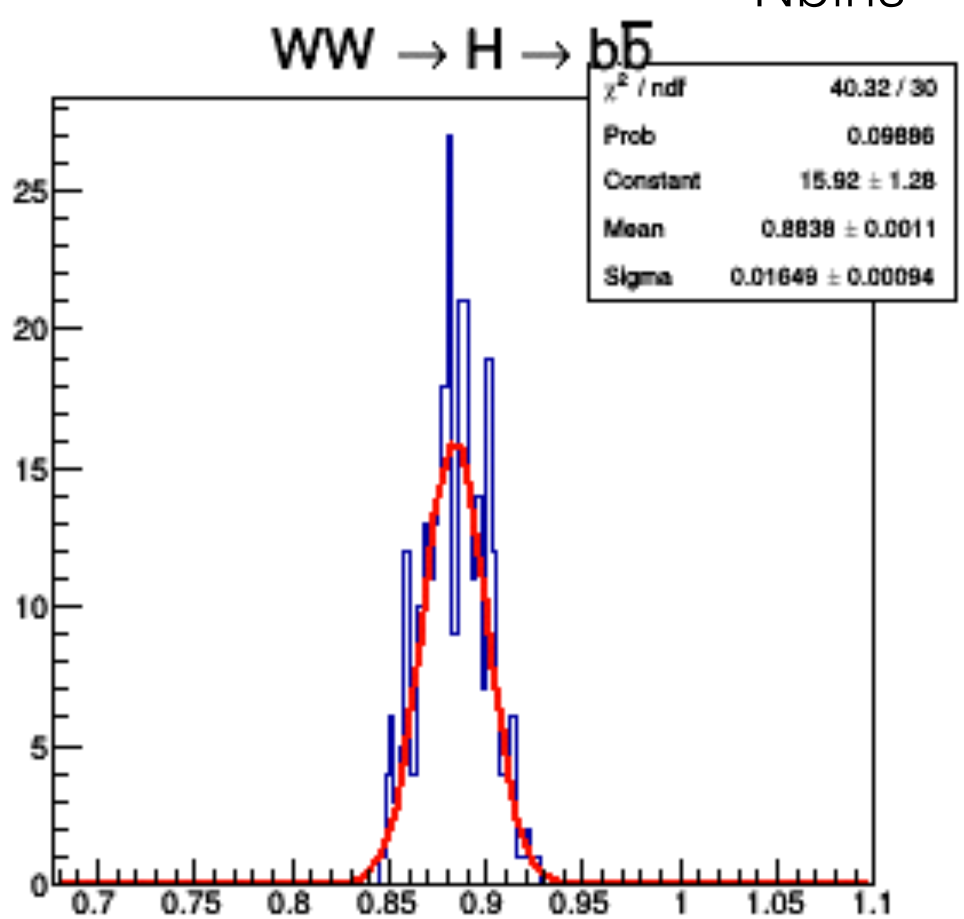


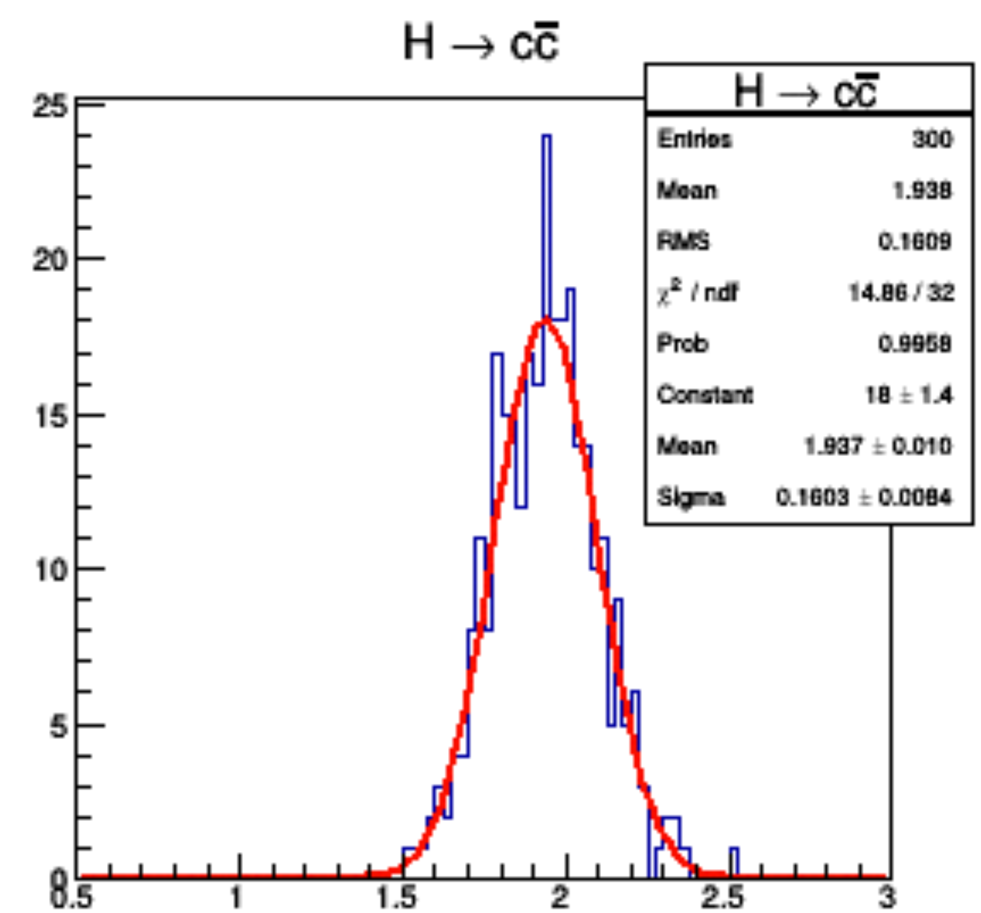


Remapping Results

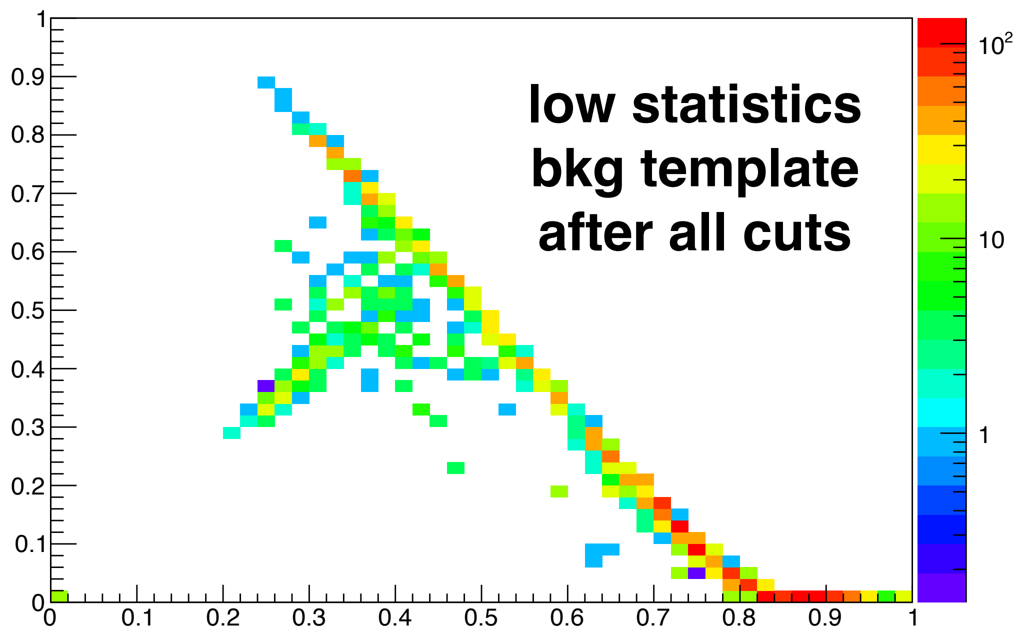


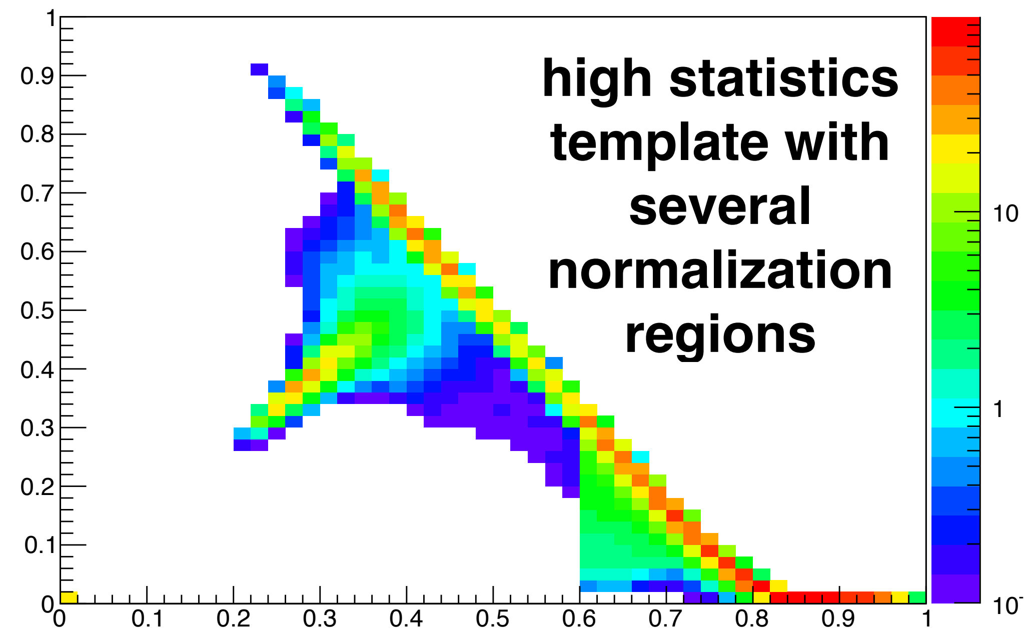






Modeling the background





- Using the "mockup" background increase the uncertainty of H branching considerably, BUT it is very difficult to segment the space and normalize since the "cliffs" are very close to the H->cc signal region
- Using dedicated Z->qq background samples to produce a high-statistics template with the right shape also reduces precision
- Smearing the histogram with smoothing functions

Conclusions

- Selection efficiency has increased drastically thanks to better training statistics for H->cc and H->gg
- Analysis now fully implements both Hvv and Hqq cutflow and performs a simultaneous fit of the templates of both
- Uncertainty on cc did not scale linearly with improved efficiency (as would be expected from background-driven data sample) The shape of the flavor space seems to have changed a lot